DOCUMENT RESUME

ED 055 362 EA 003 836

AUTHOR Guba, Egon G.
TITLE Evaluation as a Decision-Making Tool.
PUB DATE 22 Jun 70
NOTE 31p.; Speech given at Audio-Visual Conference (Indiana University, June 22, 1970)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Decision Making; *Decision Making Skills; *Evaluation; *Evaluation Methods; Speeches

ABSTRACT

This speech examines three traditional definitions of evaluation, presents a new definition, and describes how this new concept of evaluation functions. The new definition calls educational evaluation "the process of delineating, obtaining, and providing useful information for judging decision alternatives." Practical applications of this new model are presented and the model's advantage over traditional forms of evaluation are explained. (JF)

ED055362

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

# EVALUATION AS A DECISION-MAKING TOOL

Egon G. Guba

EA 003 836

Remarks Made at the Audio-Visual Conference
Indiana University
June 22, 1970

## Introduction

It is a very real pleasure for me to be here today. I had the privilege once before of addressing the Audio-Visual Conference--in June 1968--and that was an experience I thoroughly enjoyed.

I was especially pleased by the introduction today--it was the sort that my mother would believe and I would like to! As a matter of fact, it reminded me of the college football coach who was doing a bit of recruiting with this much sought-after high-school All-American. The coach said to the young man, "I understand you're quite a football player. Is that right?"

"Oh, yes," quickly came the reply, "After all, our high school team was the highest scoring team in the State, and I scored over 80 per cent of all our touchdowns."

"Very impressive," said the Coach, "What else did you do?"

"Well, I did quite a bit of running, and I averaged over seven yards a carry. And then I did all our punting--45 yards plus per kick."

"Wow, that's really something. How are your grades?"

"Straight A."

"Amazing," said the coach. "Tell me, don't you have any weaknesses?"

"Well," said the youngster, "I lie a lot!"

I am supposed to function today as keynoter, but I am afraid that to fulfill that function requires a breadth of vision that I cannot pretend to. So instead of presenting some broad overview of the field of evaluation I will focus in on one kind of evaluation that I do know

something about, a kind that sees evaluation as a handmaiden to decision-making. I shall try to make clear what such a definition of evaluation includes, and then to exemplify its application in a real setting.

Since I am going to make a somewhat different definition of the term "evaluation" than you may be accustomed to, I had perhaps better begin by giving you some more or less classic definitions that I shall specifically exclude from consideration. I shall then propose a definition which explicitly links evaluation to decision-making. In order to make the linkage clear I shall need to talk about different types of decisions and about different kinds of evaluation that service these different decision types. Finally I shall attempt to show in one continuing illustration the operational meaning of my definitions in relation to a development effort.

## Three Traditional Definitions

Evaluation, like any analytic term, can be defined in many ways. Each of the ways which have gained common acceptance has certain advantages and certain disadvantages. I should like to mention three.

1. An early definition of evaluation tended to equate that term with measurement, as it had developed in the twenties and thirties. We must remember that historically, the evaluation movement followed upon the heels of, and was made technically feasible by, the measurement movement. Moreover, the instrumentation developed by measurement experts provided the conceptual basis for evaluation. Finally, and perhaps most

important, the use of measurement devices resulted in scores and other indices that were capable of mathematical and statistical manipulation, which in turn rendered possible the handling of masses of data and the easy comparison of individual or classroom scores with group norms. Thus the idea of interpreting evaluative data in relation to an objective criterion could be introduced, but the criterion (norms) was devoid of value judgments and was, sociologically and culturally, antiseptic.

What disadvantages accrue from such a definition? First, evaluation was given an instrumental focus; the science of evaluation was viewed as the science of instrument development and interpretation. Second, the approach tended to obscure the fundamental fact that value judgments are necessarily involved (a problem to which we shall return below). Third, evaluation tended to be limited to those variables for which the science of measurement had successfully evolved instruments; other variables came to be known as "intangibles," a characterization which was equivalent to saying that they couldn't be measured; hence had no utility, and ultimately, no importance. Thus the limits placed upon evaluation because of a lack of instrumental sophistication came to be viewed as the real limits to which evaluation had to be constrained. In short, this definition results in an evaluation which is too narrow in focus and too mechanistic in its approach.

2. Another definition of evaluation which has had great currency is that of determining the congruence between performance and objectives, especially behavioral objectives. This congruence definition has had an enormous impact on education, as well it might. In the first place, the definition appeared in connection with an organized rationale about

the entire instructional process, and provided a means whereby the teacher, administrator, supervisor, and curriculum maker could make sensible judgments about what they were doing. Evaluation no longer focused solely on the student, but could provide insights about the curriculum and other educational procedures as well. The utility of evaluation was thus broadened and for the first time, a practical means was devised to provide <u>feedback</u>. Finally, evaluation came to have utility not only for judging a <u>product</u> (student achievement, for example) but also a <u>process</u> (the means of instruction, for example), a distinction whose import is only now being fully realized.

What disadvantages accrue as a result of this definition? First, with the heavy emphasis that this approach placed on <u>objectives</u>, the major task of the evaluator came to be seen as developing a set of objectives that were sufficiently operational so that the required congruence assessment could occur. The objectives themselves, in general form, were obtained by an almost mystic process that remained relatively unspecified. The real problem was to take the general objectives and by a process of successively finer definition and expansion reduce them to their most operational form.

A second disadvantage of this approach was the fact that the objectives were to be stated in <u>behavioral</u> terms. A "true" evaluation could take place only by reduction to student behaviors. Thus we are confronted with such absurdities as trying to evaluate the effectiveness of a new staff recruitment procedure, for example, by showing that this somehow related to

increased achievement on the part of students.

A third and perhaps most major disadvantage of this approach was that the emphasis on student behavior as the criterion caused evaluation to become a post facto or terminal technique. Data became available only at the end of a long instructional period. It is perhaps ironic that a definition that hinted so clearly at feedback and its utilization in improvement should have this effect. The full possibilities were thus not only not realized but the form of the definition froze evaluation as a terminal event rendering product judgments. If process data were available they could only be utilized the next time round; it was too late to use them for refinement in the ongoing program, i.e., in the program from which the evaluative data were extracted.

Thus, the definition of evaluation in congruence terms relating outcomes to objectives, while broadening the utility of evaluation considerably and providing the possibility for feedback and process data, did tend to label evaluation as a terminal process that yielded information only after the fact.

3. Neither of the two previously discussed definitions of evaluation placed much emphasis on the judgmental process. Certainly in the case of the measurement definition, and to some extent in the case of the congruence definition, the matter of placing value on the data was, if considered at all, taken pretty much for granted. But there was a school of thought that defined evaluation in yet a third way, viz., that evaluation is judgment. Perhaps the most obvious example of this definition is in the visitation procedure used by the various accrediting associations such as the North

Central Association. While evaluative criteria do exist, these are applied mainly by school personnel whose school is being evaluated, not by the visitation teams. The chief value in their application is often understood to be the process of application rather than the results obtained thereby; the school personnel through this exercise gain new insights into themselves, their problems, and their shortcomings. The actual evaluations are made not by the school personnel, however, but by the visitation teams, who come in, "soak up" the data by virtue of their expertise and experience, and render a judgment. The judgment is the evaluation.

A similar approach can be seen in the traditional school survey, and in the use of panels by the Office of Education, by Foundations, and by other funding agencies to evaluate proposals.

Advantages of this approach are fairly obvious. First, the evaluators are typically experts with a great deal of experience which they can bring into play without being artificially constrained by "instruments." Second, the evaluators are typically experts with a great deal of experience which they can bring into play without being artificially constrained by "instruments." Third, the interplay of a variety of factors in a situation is taken into account more or less automatically, and the evaluator is thus freed of the problem of relating and aggregating data after he has collected them. Finally, there is no appreciable lag between data collection and judgment; we do not need to wait for long time periods while data are being processed.

Despite these apparent advantages, however, there are very few people who would willingly rely on this approach unless nothing else can be done. First, one has the feeling that it is not so much a matter of convenience but of ignorance that forces such an approach: if we

knew more we could be more precise and objective. Secondly, we have fears for the reliability and the objectivity of such judgments, and how can one demonstrate whether they are or are not reliable and objective? It is this inability to apply the ordinary prudent tests of scientific inquiry that makes us leery, even when we are willing to concede the expertness of the evaluators involved. Third, the process hides both the data considered and the criteria or standards used to assess them, because the process is implicit. Thus, even if the judgments are valid, reliable, and objective, we have little confidence that we can tell why they are so, or to generalize to other situations. Thus, to sum up, the inherent uncertainty and ambiguity of evaluations based on this definition leave one dissatisfied.

## A New Definition: Evaluation and Decision-Making

A new definition of evaluation that I would like to discuss with you today is based on certain assumptions, viz:

1. The major task of evaluation is to service improvement in education.

2. Improvement implies change, and change implies choices from among alternative futures to which one might change.

3. The evaluator therefore does his work by servicing these choice decisions.

4. To do this evaluation must:

    a. Provide continuous readings about current status and about possible new directions.

b. Identify optional "futures" or at least discrepancies between present status and current goals.

c. Explicate values and criteria in terms of which choices will be made.

d. Provide information that weights the options in relation to the criteria.

On that basis we define evaluation as follows:

EDUCATIONAL EVALUATION IS THE /PROCESS/ OF /DELINEATING/ /OBTAINING/, AND /PROVIDING/ /USEFUL/ /INFORMATION/ FOR /JUDGING/ /DECISION ALTERNATIVES/.

This statement contains eight key terms, each of which will be found to have significant implications for the processes and techniques of evaluation. Let us take a closer look at them.

Process. A particular and continuing activity subsuming many methods and involving a number of steps or operations.

Particular attention should be paid to the fact that evaluation process is conceived as continuing; in particular, it is not conceived as terminal or as having a discrete beginning and ending. Evaluation activities are thought of as (a) sequential, i.e., with each activity forming a logical base for the next, and (b) iterative, i.e., recurrent or cyclical. These characteristics are requirements posed by the need

for continuous monitoring. Evaluation is also conceived as multifaceted, involving many different methods and techniques.

Decision alternatives. Two or more different actions that might be taken in response to some situation requiring altered action.

Educational improvement occurs only as a result of some altered action. There are at least three circumstances that might indicate that some altered action is desirable: (a) it is shown that some unmet need exists; (b) it is shown that some barrier impeding the fulfillment of a need exists (such barriers will arbitrarily be referred to as problems); or (c) it is shown that some opportunity which ought to be exploited exists. Obviously alternative needs, problems, or opportunities could be addressed. But resources are usually limited, so that some priorities must be assigned. Decisions must then be made. The alternative needs, problems, or opportunities thus constitute one class of decision alternatives; they constitute the substantial or content aspects.

But there are also formal or procedural decision alternatives. When a particular need, problem, or opportunity has been singled out for attention, there are many ways in which the need might be met, the opportunity seized, or the problem ameliorated. The several ways available must also be assessed; these ways constitute a second class of decision alternatives.

Information. Descriptive or interpretive data about entities (tangible or intangible) and their relationships, in terms of some purpose.

Webster's Dictionary defines information, among other ways, as "knowledge acquired in any manner; fact; data; learning; lore."[1] This definition is useful in reminding us that evaluation is concerned not only with scientific findings of the sort that result from research but also with data drawn from precedent and from experience. The Webster definition also serves to remind us that information can be derived in a variety of ways. It is clear that the phenomenology which the information purports to describe need not always be measureable in the rigorous sense; so-called intangibles are also eligible for inclusion when required. If conventional methods of obtaining information do not permit measurement of intangibles, it is time to extend the methodology rather than to exclude the "difficult" variables.

But information is more than a mere collection of facts and data; the facts and data must be organized to serve some purpose if they are to be intelligible. The purposes which serve as the organizational frameworks for information are typically found in the decision alternatives themselves; the information serves to differentiate the alternatives involved in the decision situation and supplies data on the basis of which the alternatives may be ordered. In this sense information may be thought of as a means for reducing the uncertainty that surrounds the decision; the more information that is available about the alternatives, the less risky the decision becomes and the better informed it is.

---

[1] Webster's New World Dictionary, College Edition, New York: World Publishing Co., 1966, p. 749.

Delineating. Identifying evaluative information required through an inventory of the decision alternatives to be weighted and the criteria to be applied in weighting them.

Evaluation is a process that furnishes information useful in guiding decision-making. Two things must be known: (a) what decision alternatives are to be considered--for it is about these alternatives that information must be obtained, and (b) what values or criteria will be applied--for the collected information must bear on these. So for example, to collect useful information relating to a decision to purchase an automobile, the evaluator must know that, say, Chevrolets, Fords, and Plymouths are to be considered, and that initial costs and economy of operation are the crucial criteria. These two sets of specifications--the range of decision alternatives and the set of criteria--can be obtained by the evaluator only in interaction with his client.

Obtaining. Making available information through such processess as collecting, organizing, and analyzing and through such formal means as measurement, data processing, and statistical analysis.

The act of obtaining will be conceived as the more technical aspect of evaluation. The evaluator as obtainer is concerned primarily (but no exclusively) with meeting the scientific criteria of evaluation such as internal and external validity, reliability, and objectivity, although the prudential criterion of efficiency is also important. To obtain implies familiarity with conventional techniques of measurement and data analysis, as well as a concern for developing methods

that meet the new demands posed by this emergent definition of evaluation. The evaluator who acts as obtainer functions in such diverse roles as instrument specialist, field data collection specialist, information system specialist, and statistician.

> Providing: Fitting information together into systems or subsystems that best serve the purposes of the evaluation and reporting the information to the decision-maker.

The act of providing involves a further interaction between the evaluator and the user of the evaluative data (the decision maker). To provide implies familiarity with the requirements of the user and with the values and criteria that are to be employed by the user, as determined during the delineation phase. It is the evaluator's function to help the client to identify his decision needs, his options, and his criteria, and then to order and highlight the evaluative data into reports that best illuminate those options within the framework of explicated criteria.

> Useful. Satisfying certain scientific, practical and prudential criteria as well as the judgmental criteria to be employed in choosing among the decision alternatives.

Information has utility (or lack of it) on two grounds, viz.: (1) it must satisfy certain criteria including the scientific criteria of internal validity, external validity, reliability, and objectivity; the practical criteria of relevance, importance, scope, credibility, timeliness,

and pervasiveness; and the prudential criterion of efficiency, and (2) it must pertain to the values and criteria which have been jointly identified by the decision-maker and the evaluator as the bases upon which the decision will be made.

Judging. The act of choosing among the several decision alternatives; the act of decision-making.

The term judging is the central term of this definition. The entire purpose of evaluation as contemplated by the definition is to service the decision-making act: to identify the decision question that calls forth an answer; to identify alternative answers (decision alternatives) that might be given in response; to identify and refine the criteria (values) to be used in choosing among available decision alternatives; to identify, collect, and report information differentiating the decision alternatives; and finally to determine whether the chosen alternative did meet expectations for it.

It is perhaps paradoxical that while the term judging is the central term of the proposed definition of evaluation, the act of judging is not central to the evaluator's role. Perhaps the clearest way to understand this distinction is to ask what would happen if a decision-maker were to engage as his own evaluator. In many ways the evaluator can be thought of as a mere extension of the decision-maker's mind; why not, in the ideal case, have a combined evaluator-decison-maker? There are no doubt decision-makers who possess the technical competence necessary to engage in the

delineating, obtaining, and providing roles briefly mentioned above, but the information provided by such a combined decison-maker-evaluator would probably not be credible to anyone else, least of all those persons most intimately affected by the decision. Similarly, the evaluator who also tried to act as decision-maker would be treated as somewhat less than completely objective. There is in short an inherent conflict of interest between the two roles that militates against their being occupied wholly or partly by the same person.

Now this conflict has an interesting corollary, for if the evaluator is totally divorced from the decison process, what prevents an unscrupulous decision-maker from making him into a dupe? Could not the decision-maker always manipulate the evaluator to his own ends by the way he defines the decision situation or names the judgmental criteria? Assuredly this possibility exists. But it seems to me that the possibility that the evaluator will be used as a dupe is less real than the almost certain probability that the evaluator will lose his objectivity if he leans too far into the decision arena. Obviously the evaluator must be alert to guard against his being caught on either horn of this dilemma.

## Types of Decisions

The particular definition which I have just explicated obviously places the decision-maker and the decisions he makes in a key role. It soon occurs to anyone who tries to apply the definition at the operational level that there are literally thousands of different decisions

that might be made in an educational setting, and that if he is to devise any kind of manageable methodology for evaluation he must somehow systematize decision-making. Unless we can find ways of grouping the many kinds of individual decisions we will have to contrive a different ad hoc evaluation design for every individual decison. Clearly that would be impractical. Thus we are confronted with the need for devising a typology or taxonomy of decisions whose categories are exhaustive of all possible educational decisions while also being mutually exclusive. Under those circumstances generalizable evaluation designs to fit all decisions that fall into similar categories becomes feasible.

I will propose a 2X2 table generated by two dimensions which I believe performs this taxonomic task adequately. Suppose we classify decisions in two ways: (1) whether they are concerned with ends or with means; and (2) whether they are concerned with intentions or with actualities. I can then assert that all educational decisions may be exhaustively and unambiguously classified as pertaining to (1) intended ends, i.e., goals, (2) intended means, i.e., procedural designs, (3) actual means, i.e., procedures in use, and (4) actual ends, i.e., attainments. This schema allows us to identify four types of educational decisions, which we shall see later can be serviced by four types of evaluation: (1) planning decisions to determine objectives, (2) structuring decisions to design procedures, (3) implementing decisions to utilize, control, and refine procedures, and (4) recycling decisions to judge or react to attainments. Let me give some example of each:

1. Planning Decisions

Planning decisions specify major changes that are needed in a program. The need for planning decisions arises from (1) awareness of a lack of agreement between what the program was intended to be and what it actually is, or (2) awareness of a lack of agreement between what the program could become and whac it is likely to become. In either case, decisions could be made to change or not to change either intentions or actualities, pertaining either to means or ends. Any such decision to introduce change would result in the establishment of program objectives.

Planning decisions are illustrated by the following questions: Should program goals be changed? Should we change or sustain our present mission? What are the top priority needs that our program should serve? What are the characteristics of the problems which must be solved in meeting the top priority needs to be served by the program? What behaviors should the students exhibit following their participation in the program?

2. Structuring Decisions

Structuring decisions specify the means to achieve the ends which have been established as a result of planning decisions. Specification of means must consider variables such as method, content, organization, personnel, schedule, facilities, and budget. Decisions about such variables arise from three sources: (1) awareness of planning decisions which specify what the program is to achieve, (2) awareness that there are alternative means available to achieve the

specified outcomes, and (3) awareness of the relative strengths and weaknesses of the available procedural alternatives. Given these three conditions, an action plan to achieve the desired objectives can be structured.

An action plan based upon structuring decisions is a comprehensive statement of outcomes to be achieved, work to be performed, and resources and time to be used. The specified outcomes are those given by the planning decisions, possibly as modified by structuring decisions in the selection of means. The decisions pertaining to work, resources, and time take the form of PERT networks, job descriptions, line-staff organizational plans, procedural specifications, process and product evaluation designs, and program budgets. Collectively, such decisions provide the operating guidelines needed to respond effectively to planning or policy decisions.

3. Implementing Decisions

Implementing decisions are those involved in carrying through the action plan. These decisions arise from two sources: (1) knowledge of the procedural specifications, and (2) continuing knowledge of the relationship between procedural specifications and actual procedures. These two kinds of information aid in process control.

Implementing decisions involve many choices regarding changes, in process, of procedures. Questions illustrating this type of decision include: Should the staff be retrained? Should new procedures be instituted? Should additional resources be sought? Should responsibilities be reassigned to staff? Should the schedule be modified? Should the public

relations activities be changed? Obviously, the making and execution of implementing decisions comprise much of the day-to-day responsibilities of operating any program.

4. Recycling Decisions

Recycling decisions are the fourth and final type of decisions in our classification schema of educational decisions. These decisions are those used in determining the relation of attainments to objectives and in determining whether to continue, terminate, evolve, or drastically modify the activity. The essential type of awareness precipitating these decisions is knowledge of the nature and timing of specified attainments.

Many questions illustrative of what we mean by recycling decisions can be posed. Are the students' needs being met? Are we solving the problems as intended? Is the program failing? Was the outcome worth the investment? Has there been a significant gain in pupil achievement? Have we benefitted by using the opportunity that was presented to us? Has sufficient progress been achieved to warrant continuation of the program? Is the new program succeeding? Were the results from Program A better than those from Program B? Was the procedure effective? Has the program resulted in improved teacher competence? Have school-community relations been improved? Have students improved their self-concepts? Questions such as these often must be answered when operations managers are attempting to justify new funding requests. Continuing to fund expensive procedures without

answering such questions understandably is often frowned on by responsible fiscal agents.

## Types of Evaluation

Corresponding to each of these four decision types are four types of evaluation, which might be thought of as four generalizable evaluation designs; we shall give the four types the names *context*, *input, process, and product*. It might be noted that the initial letters of these four terms from the acronym CIPP (pronounced *sip*) which is often used as a general name for the formulations propounded here. *Context* evaluation services planning decisions, *input* evaluation services *structuring decision*, *process* evaluation services *implementing decisions, product* evaluation services *recycling decisions*. We shall discuss each in turn.

*Context evaluation* services planning decisions. Its major objective is to define the environment about which decisions are being made, to depict unmet needs, to identify problems that prevent needs from being met, and to identify opportunities that should be seized. It is a continuous process that presents data to the decision-maker at frequent intervals. Its purpose is at least as much to *create awareness* of the need for a decision as it is to delimit the domain of that decision. Context evaluation may focus on the inward workings of the decision-maker's agency, in which case it may be viewed as a kind of process control mechanism, and/or it may focus on the outside environment

to take advantage of new contingencies or opportunities. It identifies needs to be met and problems to be solved, and furnishes information about their priorities. Context evaluation thus creates an awareness in the decision-maker that he must make a planning decision and furnishes him a context of information within which to make it.

Input evaluation services structuring decisions. Needs or problems illuminated by context evaluation require some response. This response may, on reflection, take the form of enlightened persistence (which, in its more perverse form, becomes maintaining the status quo) or of informed action. Given that some need or problem has been identified in relation to which action is proposed, the objective of input evaluation becomes that of identifying and assessing relevant capabilities of the action agency, strategies which may be appropriate for meeting program goals, and tactics (designs) that are appropriate to the selected strategy. Input evaluation thus produces an analysis of alternative procedural designs in terms of potential costs and benefits. It is not a continuous process but evolves ad hoc after an appropriate planning decision has been made.

Process evaluation services implementing decisions. Once a designed course of action has been approved and implementation of the design has begun, process evaluation is needed to provide periodic feedback to the decision-maker responsible for continuous control and refinement of plans and procedures. The objective of process evaluation is to detect or predict, during the implementation stages, defects in the procedural

design or in its implementation. Like input evaluation, it is ad hoc in nature, being called into play only when there is a particular procedural design to be evluated. Process evaluation creates in the decision-maker either an awareness that a refinement is needed or gives him reassurance that all is well.

Product evaluation services recycling decisions. The objective of product evaluation is to measure and interpret attainments, not only at the end of a project cycle but as often as necessary during the duration of the project. Product evaluation provides information for deciding whether to continue, to recycle, to modify, or to terminate the activity which is being evaluated. Like input and process evaluation, it is ad hoc in nature. Product evaluation assures the decision-maker that a proposed action is resulting in outcomes planned for, or provides him evidence about the ways in which it is falling short.

## An Example

I would like now to illustrate these theoretical statements with an extended example. Suppose that a particular development agency is concerned with providing better educational opportunities for the children of agricultural migrants. Let us see what such an agency might do given an adequate evaluation approach along the lines I have suggested.

As a first step it is important for the agency to identify the needs, problems, and opportunities that beset or typify this particular audience and to choose from among those identified needs, problems, and

opportunities that subset (which may be a subset of one) to which it will attempt to respond. The development agency begins by depicting the domain which it is called upon to service. It identifies the boundaries of that domain by defining what will be taken to the population of migrant children, using the definition in the federal law that made the funds available in the first place. Various system elements will then be defined about which data or information must be collected; thus the children themselves, the schools in which they are located (or pass through), their parents, the schools programs, their teachers and other related educational personnel, and the like will be named. The characteristics of each such element will then be defined, e.g., in the case of the children themselves, such factors as age, sex, IQ, placement, income level, sibling order, and the like may all be important. Such factors serve to depict the domain, and they will be systematically surveyed as part of the context evaluation.

As next step the agency will identify need, problem, and opportunity candidates, i.e., the needs of the target population to which they might respond, the problems that prevent those needs from being otherwise fulfilled, and the opportunities that exist for serving this audience. For the hypothetical target population of migrants such needs might include the need for more occupational information (in view of the fact that migrant labor needs are diminishing and new occupational outlets will probably be necessary for migrant youngsters), and the need for better health care, the need for proper nourishment, the need for a supportive home environment, and the like.

The problems of this target population include mobility (which in turn induces such problems as program discontinuities, teacher contact discontinuities, and lack of closure), retardation, language difficulties (since most migrant farm youngsters are Spanish-American and speak only that language upon first coming to school), cultural differences, dysfunctional personality characteristics, high drop-out rate, and dysfunctional school responses to their plight. Opportunities include the availability of federal funds in support of the target population.

A third step in the context evaluation (which may go on simultaneously with the other two) is the identification of criteria in terms of which the decision among need, problem, or opportunity candidates will be based. This identification is necessary to guide the context evaluator in collecting appropriate information about the alternative needs, problems or opportunities. In order to gather these criterion data the evaluator must work in a face-to-face relationship with the decision-maker and often must help him to make explicit those criteria which were heretofore only implicit in his mind.

The particular criteria in any real case will of course vary widely from agency to agency and audience to audience. In the example we might imagine that they would include such as cost, personnel requirements, time requirements, probable benefits, probable side effects, possible relationships (building upon or providing inputs) to other agency programs, political viability, social viability, and the like. For each criterion identified the context evaluation should provide appropriate data. Thus, in the case of cost, some estimate should be made of the cost of responding

to any given need, problem, or opportunity, as one basis for aiding the decision about which one to service.

A final step in the context evaluation is working with the decision-maker to decide which needs, problems, or opportunities are to be serviced. This is not a simple matter of displaying all possible alternatives and evaluating each on the criteria identified. As a matter of sheer logistics not all possible alternatives could have been identified anyway. Nor will all criteria have been identified beforehand; some remain as "hidden agenda" items and others will evolve only from the interaction of the decision-maker with the evaluative date provided. The evaluator's task is thus not just one of transmitting codified information but of working with the decision-maker to insure its productive use.

We may assume, then, that when we have reached the termination of the context evaluation phase, the decision-maker will have selected the needs, problems, or opportunities to be serviced and that the context date will provide the basis for delineating the specifications which any proposed response must meet. The ends of the agency's development activity will now be clearly formulated. We shall refer to this statement of ends as the "ends specifications."

Let us assume that a problem has been selected and that specifically, the problem of mobility will be dealt with. It may be argued that whatever the educational problems of this group may be, they are enormously intensified by the fact that these children move so frequently. The agency thus sets as its end the development of a device for dealing with this problem. We are then ready to move into the input evaluation phase, and evaluation

which is, it must be noted, ad hoc to this problem. The context evaluation was not constrained to identify any particular need, problem, or opportunity; the input evaluation is constrained to identify a solution to the particular problem which has been previously selected for attack, i.e., migrant mobility.

Input evaluation may itself be thought of as having two phases: that related to strategy selection and that related to tactics selection. Let us pursue the migrant example further. If mobility is the problem to be dealt with, there are obviously a number of strategies that might be employed. Thus, it might be proposed that the legislature pass a law forbidding school age children to migrate and requiring their continuous attendance in one school district. Or, it might be proposed to devise mobile classrooms (trailers) staffed with appropriately trained teachers who would follow the migrant streams and thus continuously relate to the same children. It might be proposed to devise individualized instructional packages, properly programmed, which each child might carry with him wherever he went to school.

The decision as to which strategy should be pursued can be served in a variety of ways. Expert opinion might be solicited concerning the viability of any proposed strategy; thus we might soon find, by asking political figures, that the strategy of passing an appropriate law is simply not viable. This is especially true because of the ethnic composition of the target audience. Certain of the descriptive data collected during the context period would quickly invalidate the mobile classroom idea; migrants simply do not travel in well ordered groups making it impossible

for a teacher to follow the same children in any systematic way. It might also be possible to study existing examples which have already built up some experience; thus a visit to a school using the Individually Prescribed Instructional Materials (IPI) developed by the Pittsburgh Research and Development Center might produce some evidence in favor of the individualized approach. All such data collection is properly evaluation in the input sense.

The second phase of input evaluation has to do with the development of the tactics necessary to implement a selected strategy. Let us assume that the decision has now been made to pursue the strategy of developing individualized materials. What shall such materials be like? Shall they involve a variety of subject matter or only tool skills? Shall they be programmed materials or textually arranged? Shall one think in terms of films and filmstrips or only of printed materials? Is one format better than the other? How can one arrange for pupil reinforcement? How can one arrange to get pupils' questions answered? Dealing with matters such as these is of the essence for the developer; it is how developments are, in fact, engineered. It is the evaluator's task to provide the information necessary for these decisions, and if the information is not readily available to arrange for appropriate studies to get it.

It should be noted that both development and input evaluation are essentially in-house activities. While there may be contact with the real world for certain purposes, as for example, to determine whether Format A or Format B give better results, these contacts are essentially controlled by the evaluator for his purposes. These controlled contacts will be referred

to in this paper as pilot test activities (not to be confused with field tests which will be discussed below): the purpose of pilot tests is to determine whether components of the overall strategy (the individual tactics) perform more or less up to expectations under controlled conditions, in much the same manner that a new carburetor might be bench-tested before being installed in a real auto for a real world test. The ends specifications evolved during the context evaluation serve as the ultimate criterion for these tests, but the satisfaction of ends specifications under laboratory conditions cannot of course be taken as absolute evidence for their satisfactory performance in the real world. Nevertheless it would be irrational to assemble a prototype without some assurance that the parts conform to design requirements.

The end product of the development process, aided and abetted by the input evaluation, is thus a working prototype of the response to the problem. The prototype components are reasonably in conformity with the ends specifications resulting from the context evaluation. The development process augments the ends specifications with a second set of specifications, which we shall term the "means specifications," which indicate how the prototype is expected to be installed and operated. We are now ready to take the prototype from the antiseptic development laboratory and insert it into the septic world; we are ready for process and product evaluation, or field tests.

Process and product evaluation are thus also ad hoc to a particular prototype which has been evolved. Process evaluation is concerned with whether or not means specifications are satisfied while product evaluation

is concerned with whether or not ends specifications are satisfied. Both go on in the real world, and both go on simultaneously although not necessarily with equal emphasis at all times.

Initially the emphasis--amount of effort--is placed on process evaluation. The first concern must be with whether the prototype is installed and working as one expects. There is always a good deal of "debugging" that must go on when a change is introduced; this debugging is in the province of process evaluation. Are the teachers teaching as they should? Do the materials arrive on time? Is the sequencing clear? Are the projected resources sufficient?

Product evaluation is likely to receive heavy emphasis once the debugging is complete and the process seems to be going well. Then we are likely to begin asking questions like: Are the students learning? Have they progressed to where we thought they ought to be at this point? Should we continue with the cycle or are refinements of some kind necessary? When it is important to do so, as for example, because we contemplate using the developed solution in a variety of settings other than the one in which it is being tested, we may wish to use experimental approaches (once process evaluation indicates that procedurally things are on good order) that will allow wide generalizability. In such cases we may wish to declare a moratorium on further refinements. More typically, however, we will wish to use both process and product data to produce continuous refinements and improvements both in substance and procedures.

In the case of the migrant example, both process and product

evaluation take place in the real world of the migrant child. In terms of process evaluation, we would be asking questions such as: Does each child receive the appropriate materials on time? Is he adequately instructed in their use? Does he remember to carry them along? Does he lose them? Can he use them, e.g., can he actually view a filmstrip with which he has been provided or does he find that he cannot plug in the projector because there is no power source? Do teachers in transient schools know how to pick up with each migrant child and relate to his semi-completed work? Can grading and credit systems be adopted to the program? And the like.

In the case of product evaluation, we would be asking questions like: Are program discontinuities in fact eliminated by this approach? Do children learn at a rate comparable to the rate that might be expected of them if they were permanent residents somewhere? Are teachers effective in the new roles they must play?

In general, process evaluation would relate to the satisfaction of means specifications while product evaluation would relate to the satisfaction of ends specifications. Data in both cases are collected continuously, and both kinds of data can be used to refine, improve, recycle, confirm, or discontinue a program at any time. Both are carried on in the field under real world conditions (that is, under conditions of "invited interference" rather than under the controlled conditions that typify the laboratory). Process evaluation receives initial emphasis but is never discontinued; it becomes, in the end, a kind of process control. Product evaluation receives later emphasis but it too is continuous.

If the prototype solution survives this field test it is ready for permanent installation. If the development agency has continued to be concerned with the target audience throughout this period, then the context evaluation machanism which was initially developed is still functioning. The newly installed prototype then comes under the purview of this context evaluation mechanism, whose information will preseumably now show that the original problem has been eliminated or ameliorated. A new need, problem, or opportunity may now achieve top priority, and the whole process is started again. Or, conversely, there may no longer be a need, problem, or opportunity of sufficient magnitude to which to respond, so that a policy of "enlightened persistence" is counseled. In either case the context mechanism can be programed to continue its regular probing, creating an awareness that another planning decision is necessary should that contingency arise.

## Finale

Well, I have prattled on for an unconscionably long time. What I have tried to do is to get you thinking about a new model of evaluation. You must understand that the new model is by no means thoroughly explicated, nor are the wide variety of techniques, instruments, and processes that are necessary to its full application available. I would not delude you into thinking that it is easier, cheaper, or more efficient than are other formulations of evaluation. But I believe that it has real advantages over existing formulations, which I have tried to illustrate. I await your questions and comments to see whether I have been at all successful.